Business Presentation

Presenter : Esteban Ordenes

Contents

Business Problem Overview and Solution Approach

- Travel Package Purchase Prediction

- Core business idea
  - Currently, there are 5 types of packages the company is offering - Basic, Standard, Deluxe, Super Deluxe, King. Looking

    at the data of the last year, we observed that 18% of the customers purchased the packages. However, the marketing cost

    was quite high because customers were contacted at random without looking at the available information. The company is

    now planning to launch a new product i.e. Wellness Tourism Package. Wellness Tourism is defined as Travel that allows

    the traveler to maintain, enhance or kick-start a healthy lifestyle, and support or increase one's sense of well-being.

    However, this time company wants to harness the available data of existing and potential customers to make the

    marketing expenditure more efficient.

- Problem to tackle
  - Build a model that will help the marketing department to identify the potential customers who have a higher
    probability of purchasing the Tourist Package.

# Data Overview

- **Data Dictionary**
- CustomerID: Unique customer ID
- ProdTaken: Whether the customer has purchased a package or not (0: No, 1: Yes)
- Age: Age of customer
- TypeofContact: How customer was contacted (Company Invited or Self Inquiry)
- CityTier: City tier depends on the development of a city, population, facilities, and living standards.
- Occupation: Occupation of customer
- Gender: Gender of customer
- NumberOfPersonVisiting: Total number of persons planning to take the trip with the customer
- PreferredPropertyStar: Preferred hotel property rating by customer
- MaritalStatus: Marital status of customer
- NumberOfTrips: Average number of trips in a year by customer
- Passport: The customer has a passport or not (0: No, 1: Yes)
- OwnCar: Whether the customers own a car or not (0: No, 1: Yes)
- NumberOfChildrenVisiting: Total number of children with age less than 5 planning to take the trip with the customer
- Designation: Designation of the customer in the current organization
- MonthlyIncome: Gross monthly income of the customer
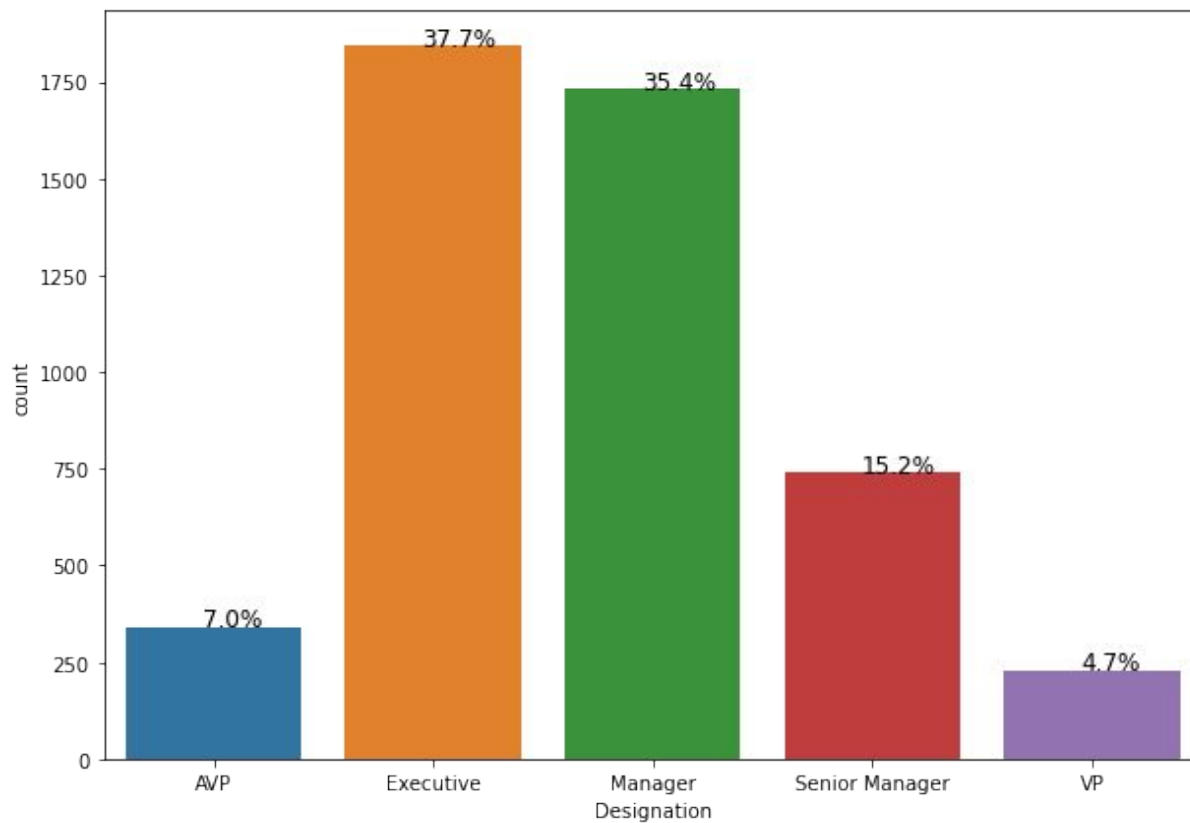
Data Overview

- **Customer interaction data**
- PitchSatisfactionScore: Sales pitch satisfaction score
- ProductPitched: Product pitched by the salesperson
- NumberOfFollowups: Total number of follow-ups has been done by the salesperson after the sales pitch
- DurationOfPitch: Duration of the pitch by a salesperson to the customer
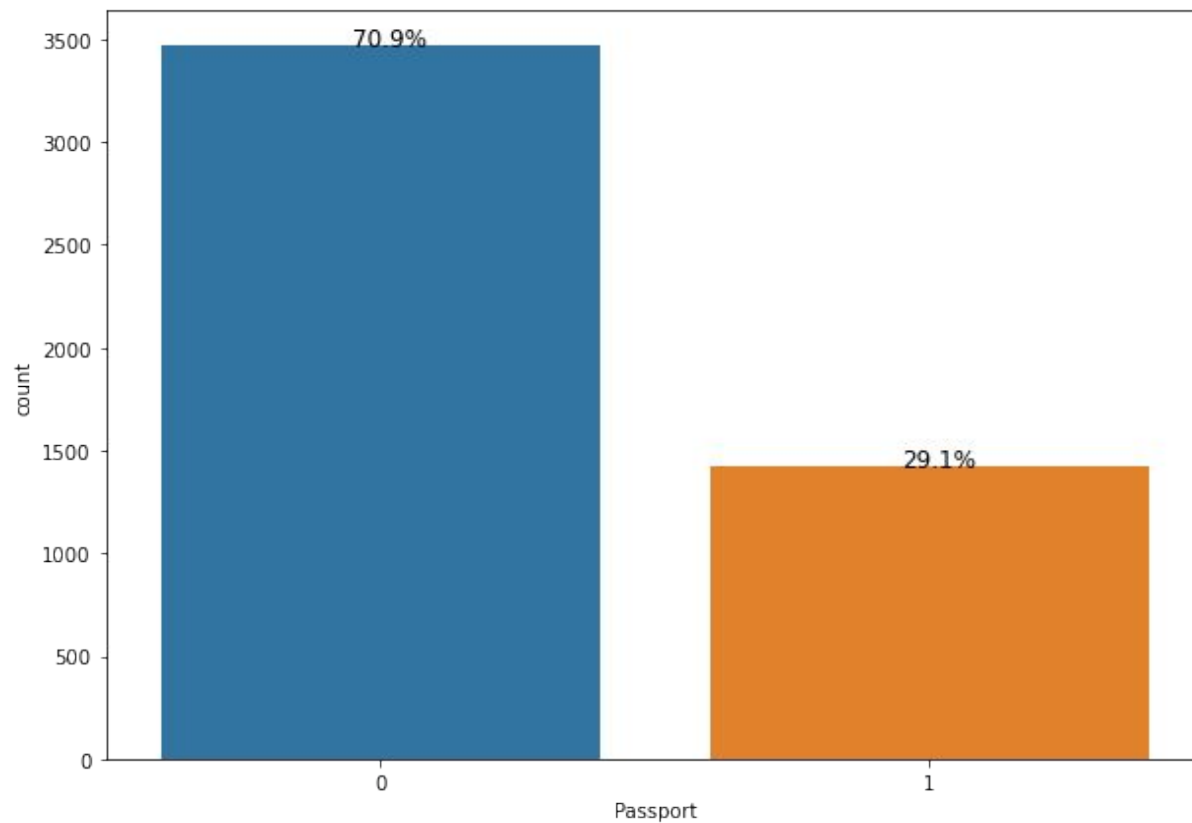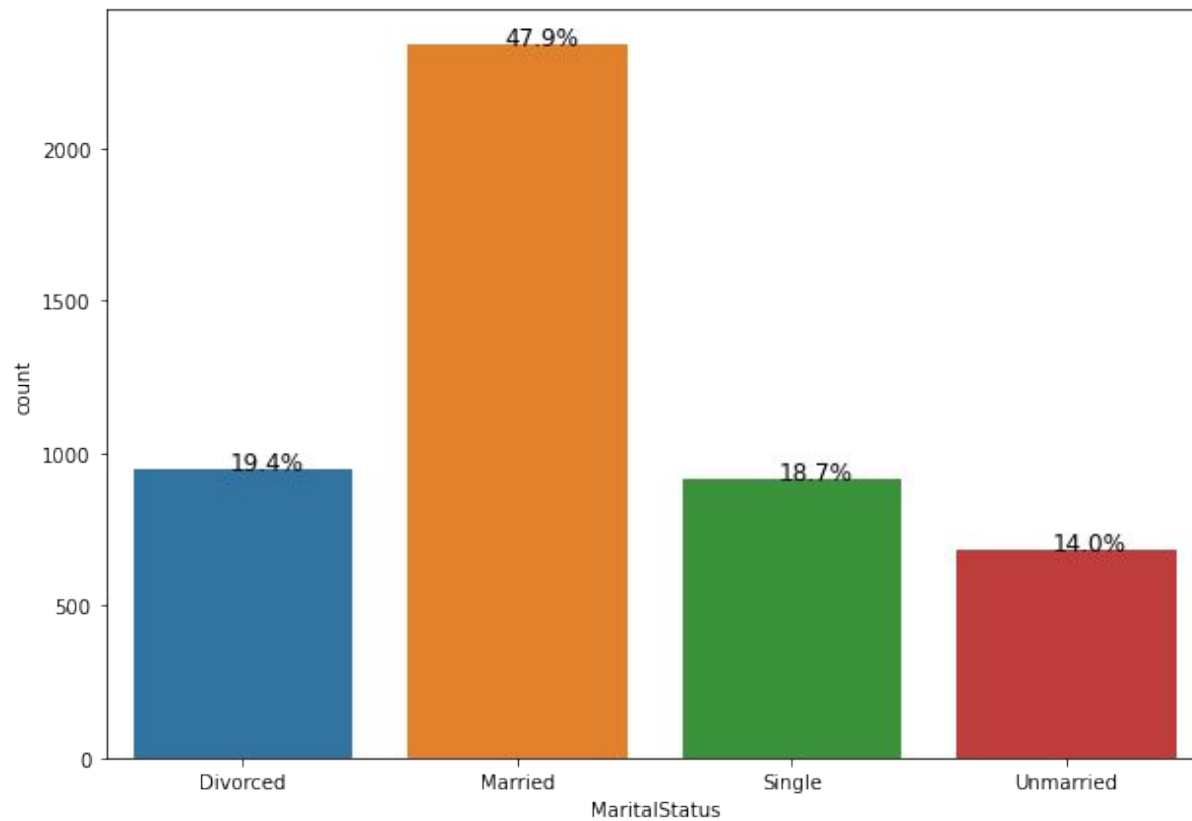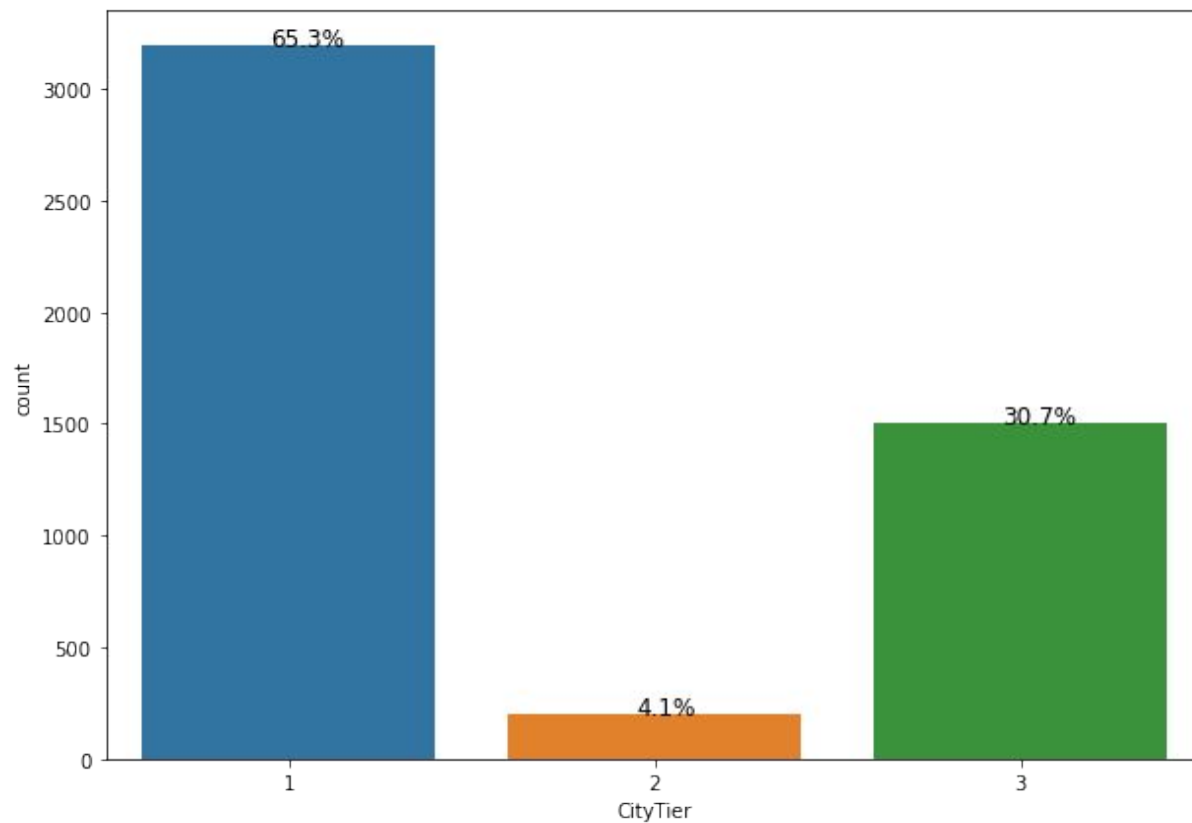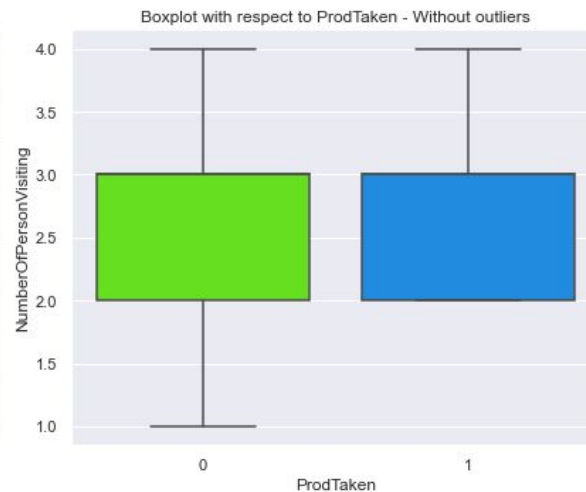
EDA

EDA

EDA

EDA

EDA

EDA
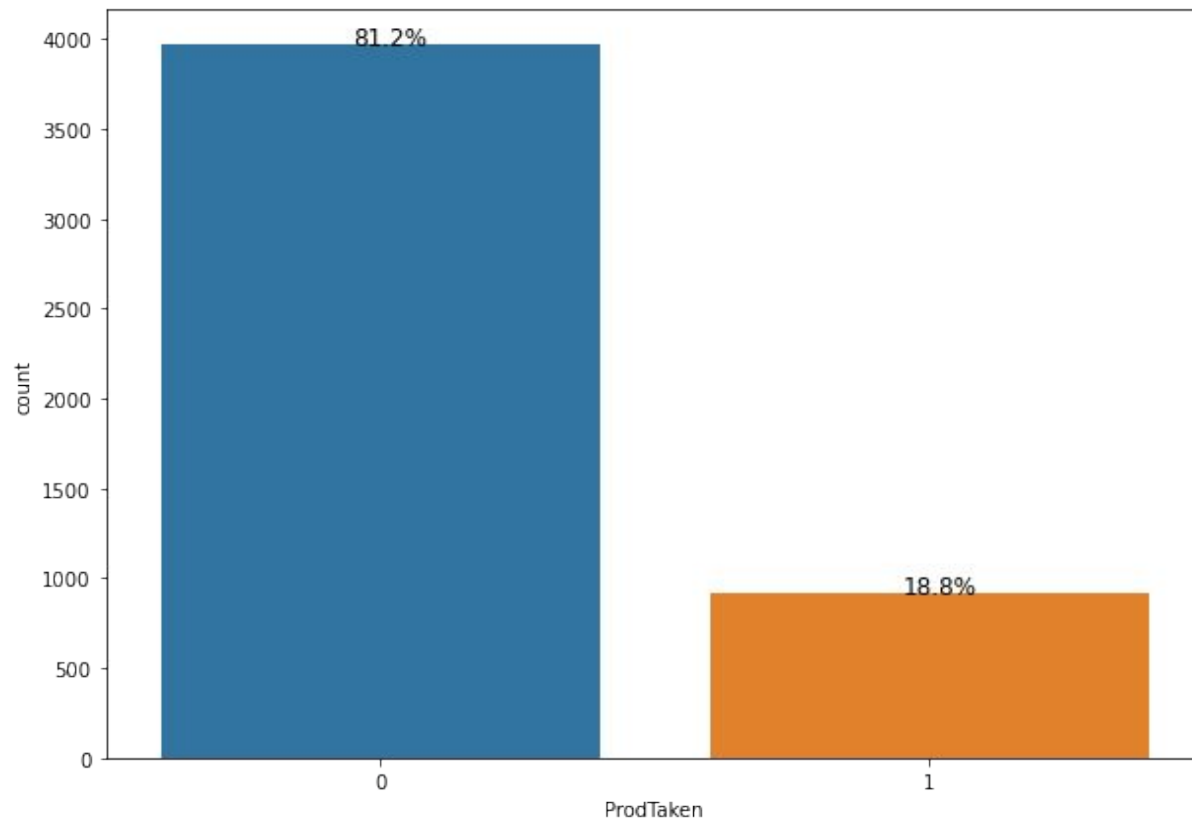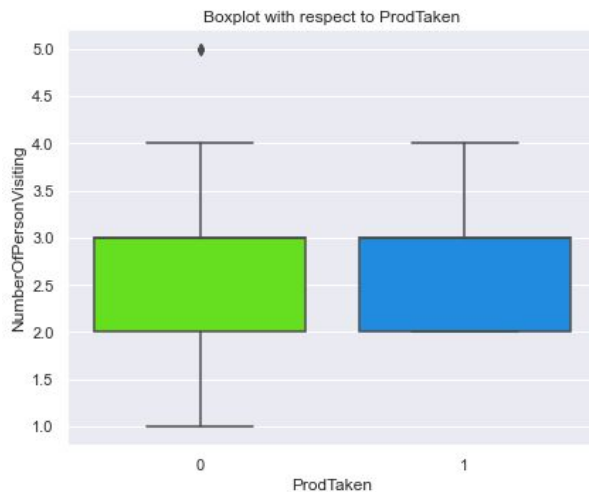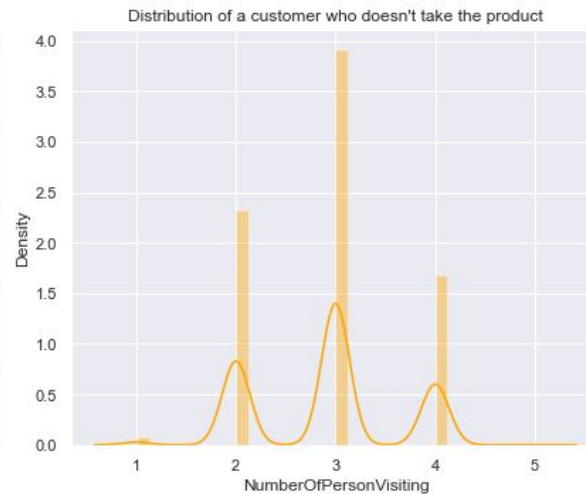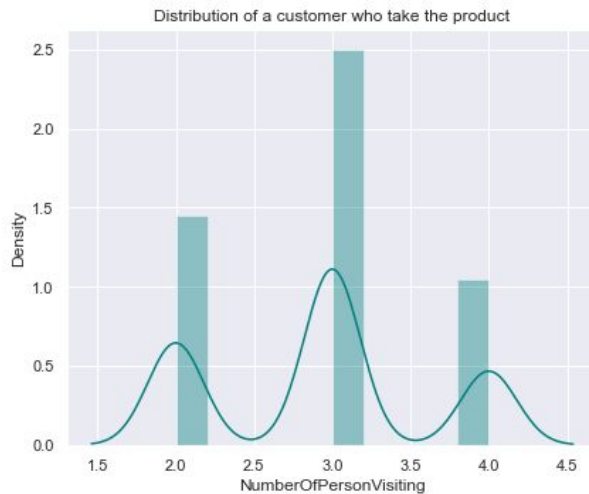
EDA
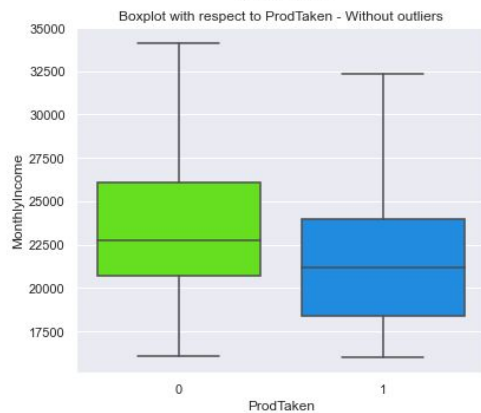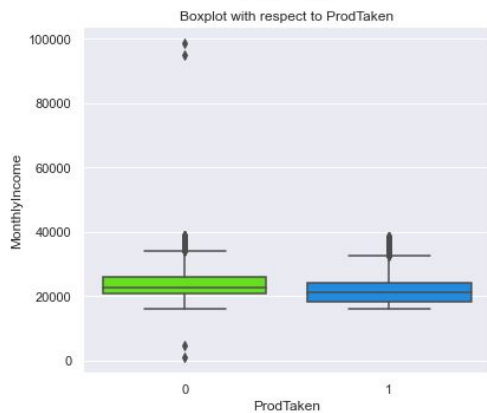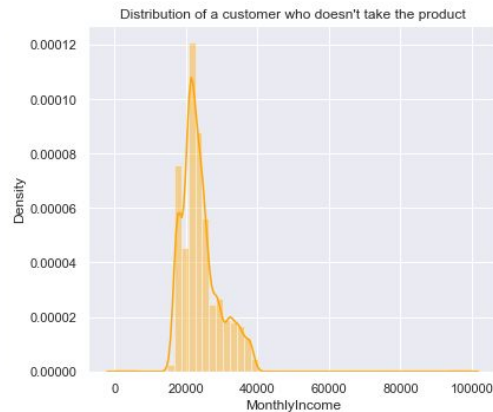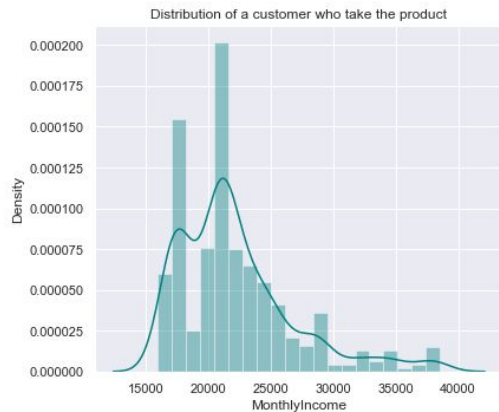
EDA
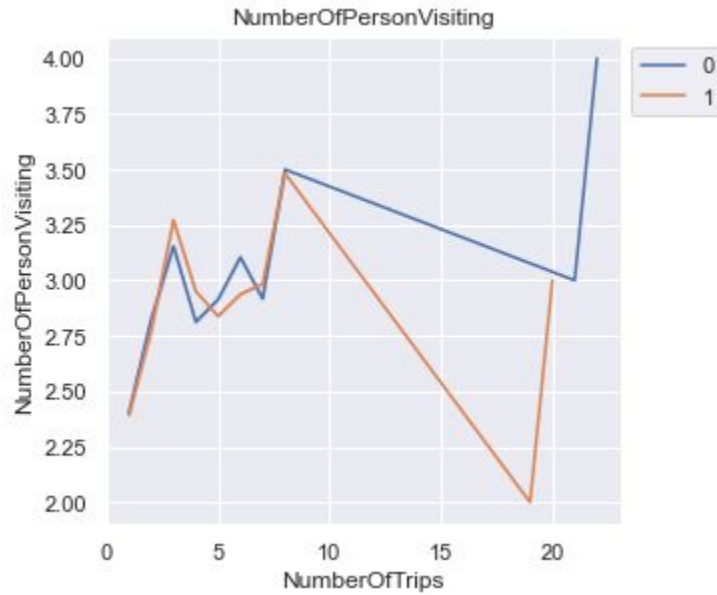
EDA

EDA

EDA

Bagging Model Performance Summary

Bagging Model Performance Summary

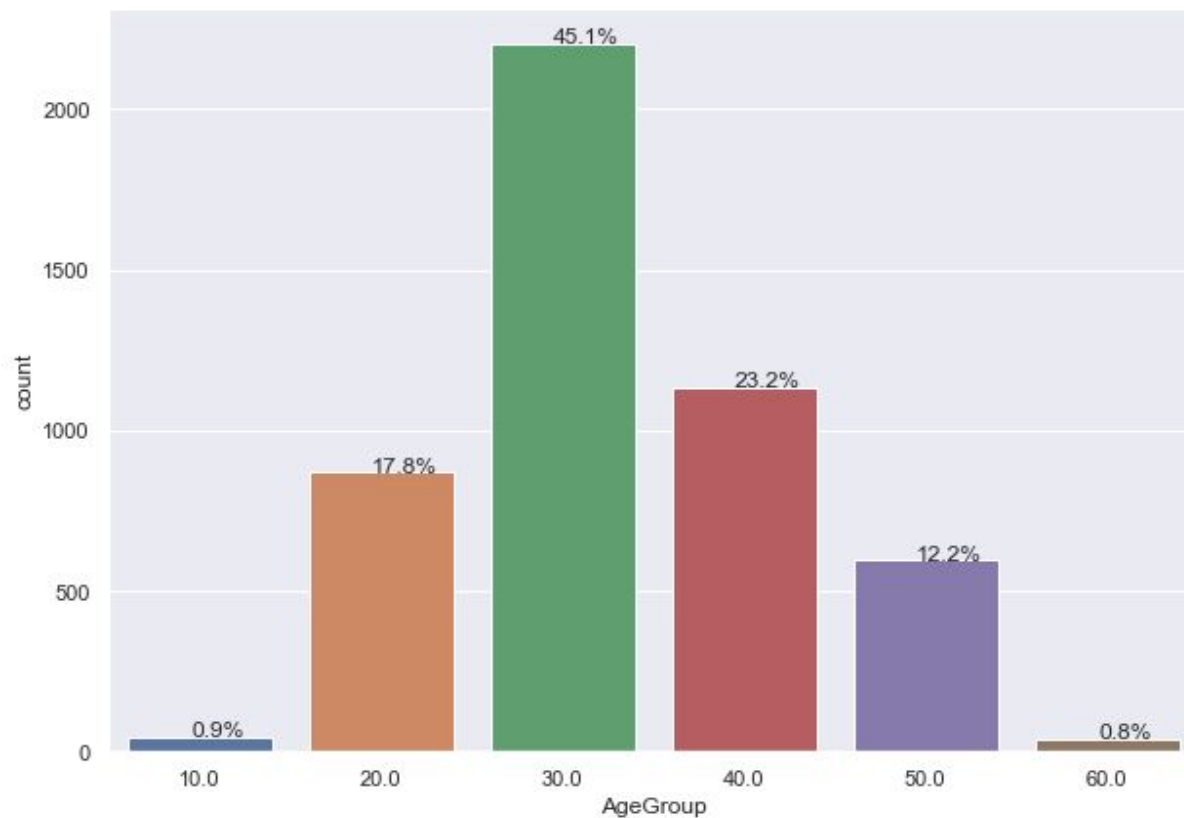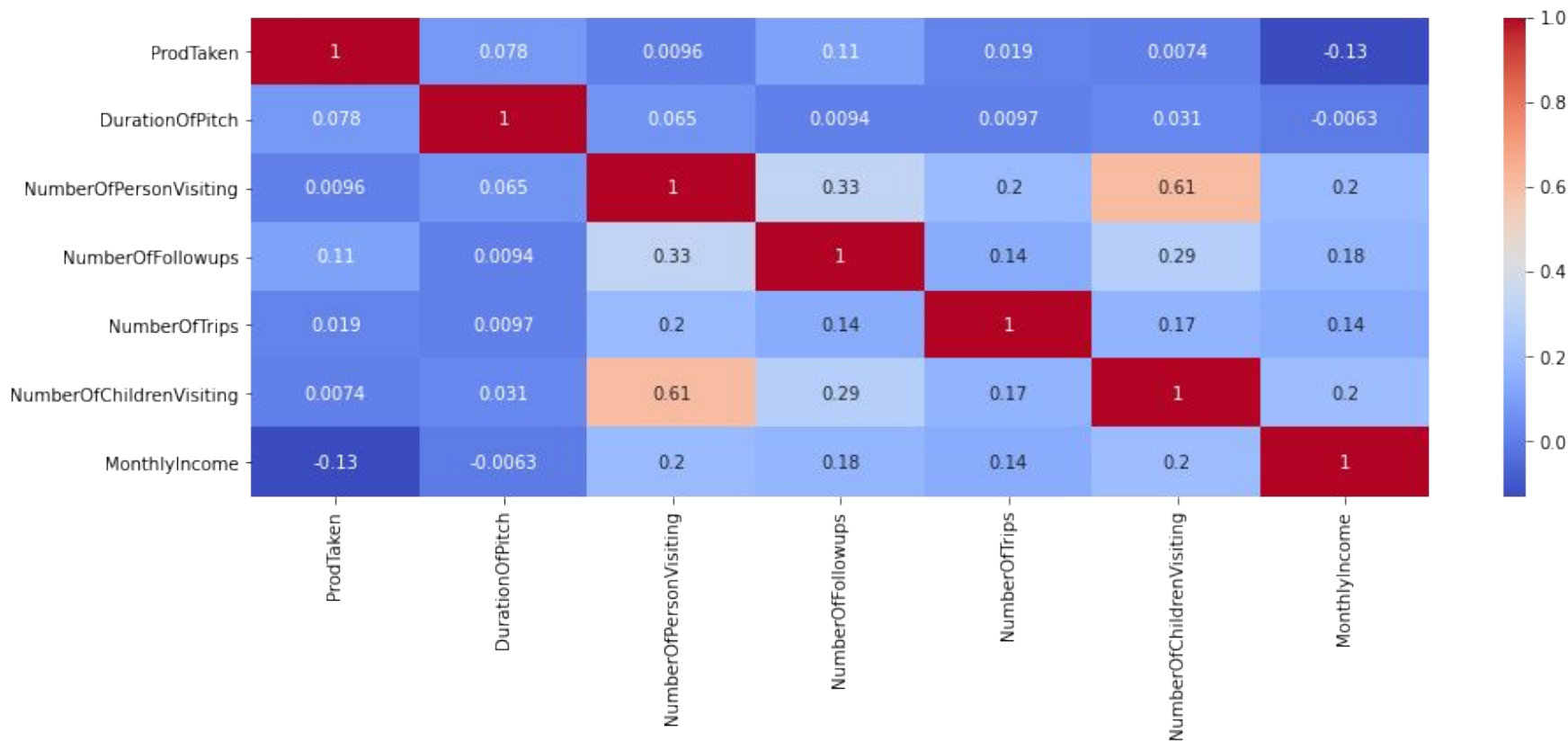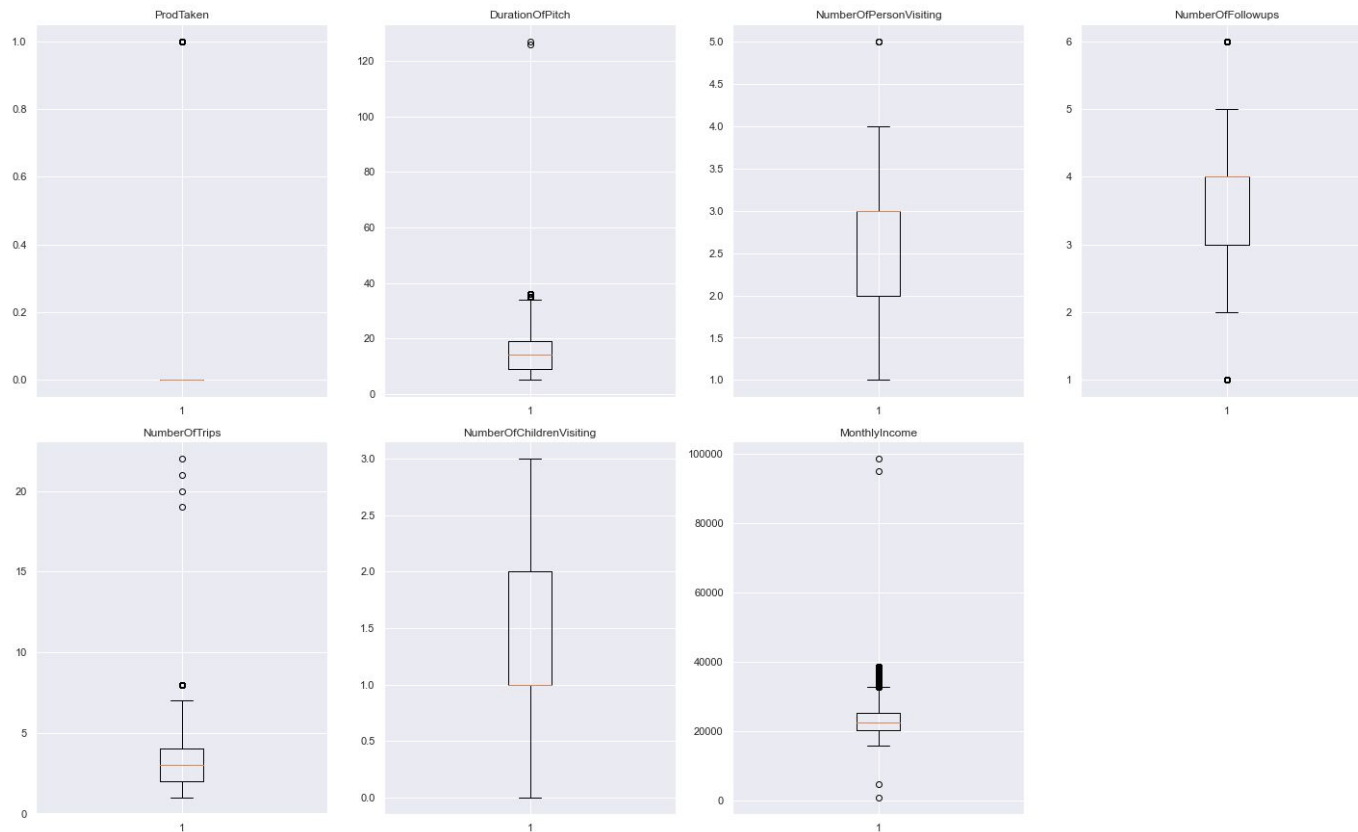|	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision
0	Decision Tree	1.000000	0.890934	1.000000	0.731884	1.000000	0.701389
1	Tuned Decision Tree	0.188249	0.188139	1.000000	1.000000	0.188249	0.188139
2	Bagging Classifier	0.994446	0.910702	0.972050	0.644928	0.998405	0.843602
3	Weighted Bagging Classifier	0.995615	0.903885	0.978261	0.565217	0.998415	0.881356
4	Tuned Bagging Classifier	0.481146	0.422631	0.961180	0.891304	0.261292	0.231421
5	Random Forest	1.000000	0.914110	1.000000	0.579710	1.000000	0.941176
6	Weighted Random Forest	1.000000	0.901840	1.000000	0.525362	1.000000	0.917722
7	Tuned Random Forest	1.000000	0.929107	1.000000	0.702899	1.000000	0.898148

- Decision tree performed well on training and test set.
- Bagging classifier overfitted the data before and after tuning.
- Random Forest with default parameters performed better after tuning.

Bagging Model Performance Summary
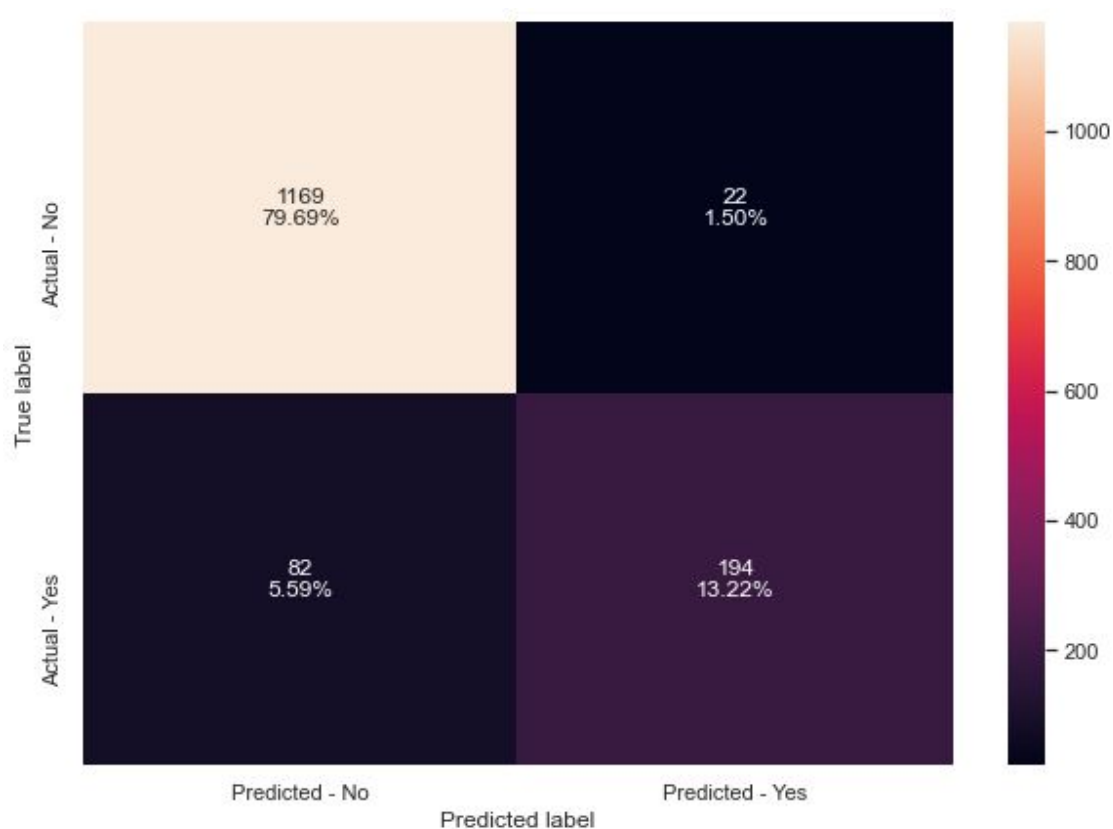

Feature Importances
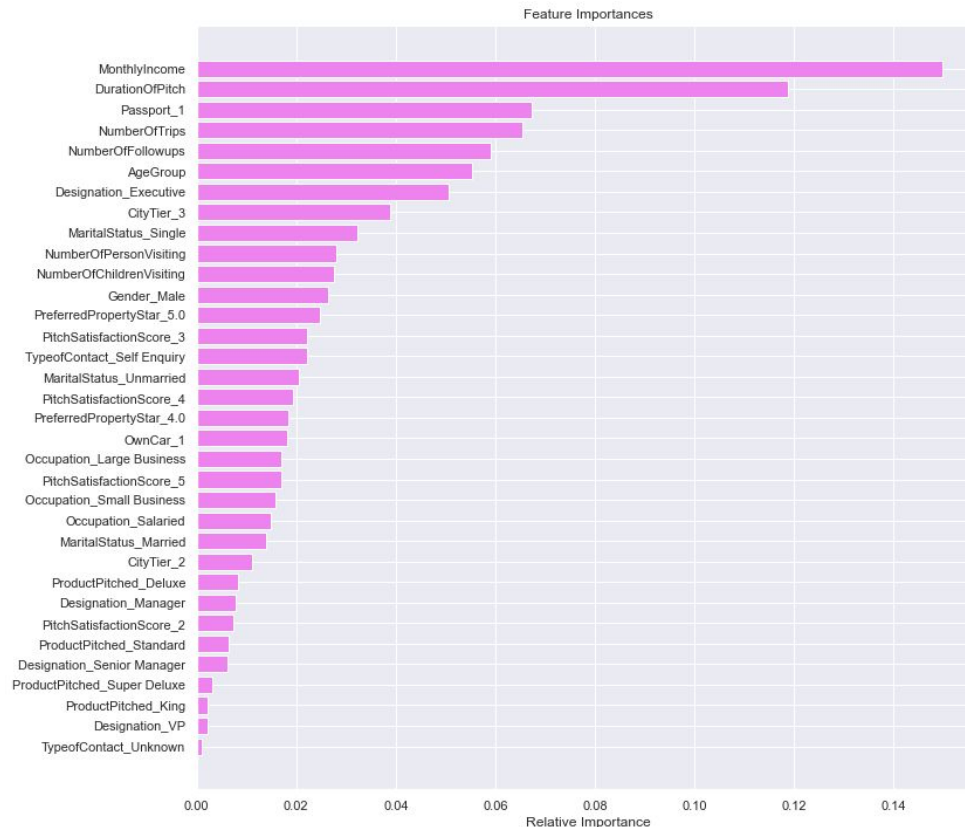
Boosting Model Performance Summary

Boosting Model Performance Summary
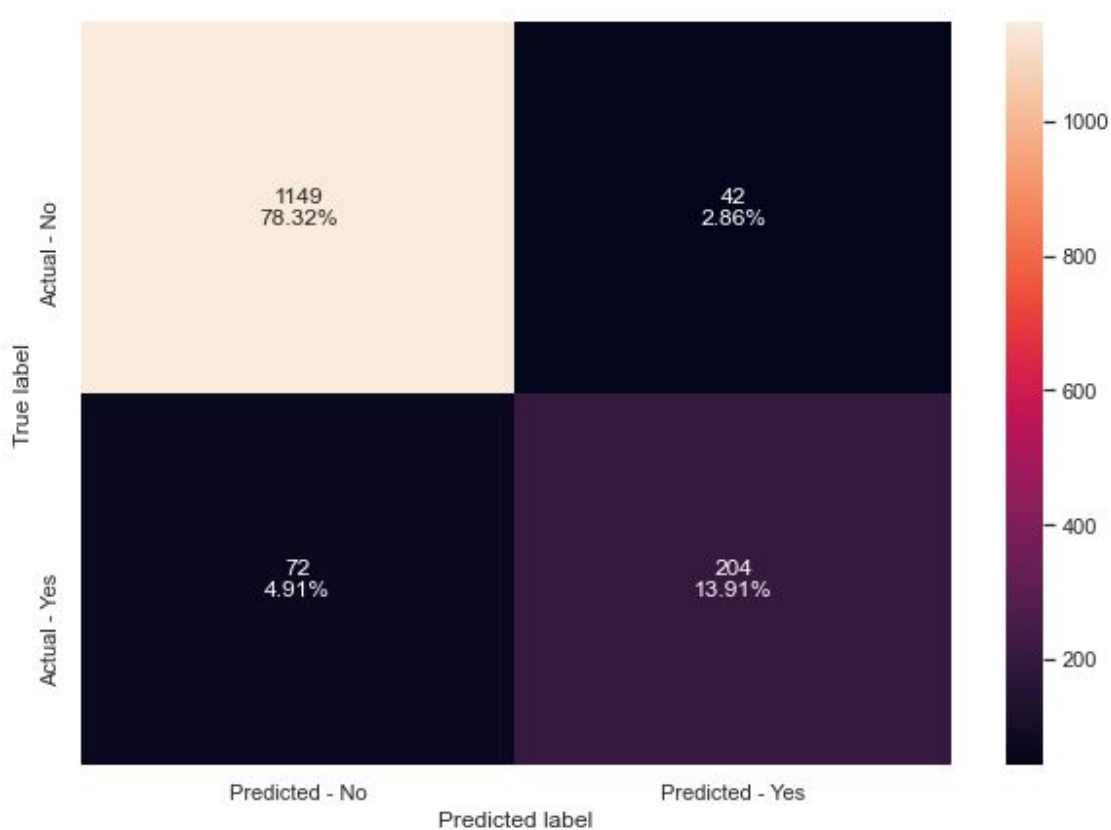
	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision
0	AdaBoost with default paramters	0.84	0.84	0.31	0.30	0.68	0.68
1	AdaBoost Tuned	0.99	0.88	0.95	0.62	0.98	0.69
2	Gradient Boosting with init=AdaBoost	0.88	0.87	0.44	0.40	0.87	0.79
3	Gradient Boosting Tuned	0.91	0.88	0.56	0.49	0.92	0.77
4	XGBoost Tuned	0.99	0.92	0.98	0.74	0.98	0.83

- Tuned XGBoost model is the best model here. It has really high performance metrics, and consistent recall values.

Boosting Model Performance Summary

Feature Importances
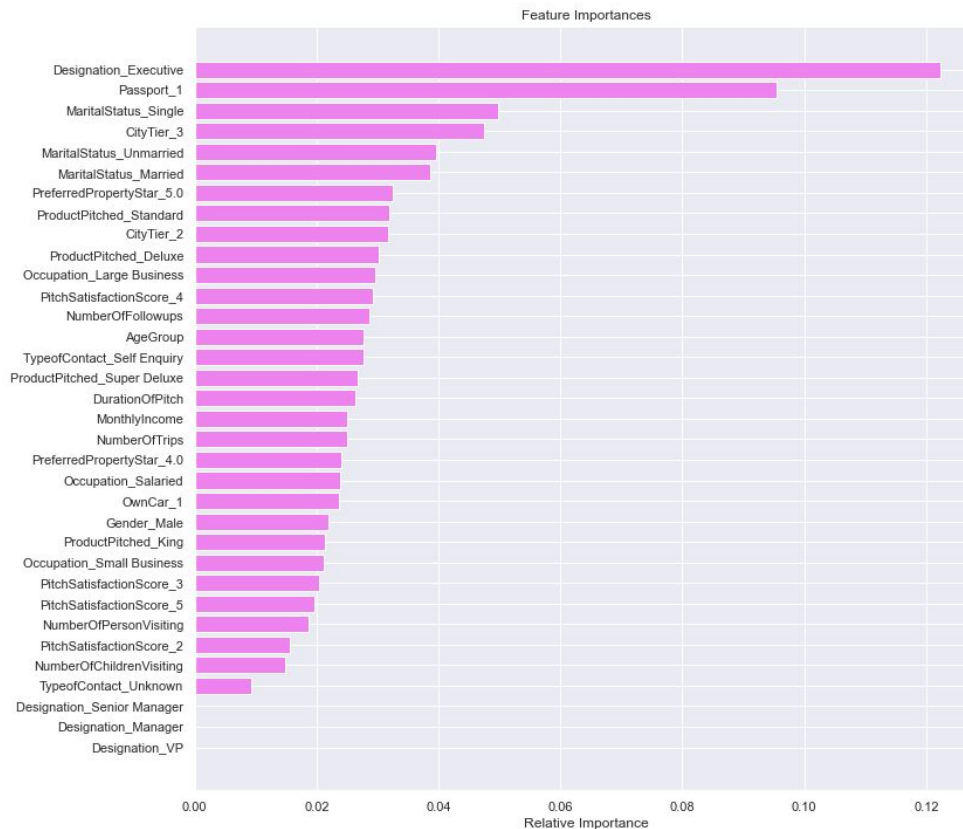
Conclusion

- Tuned Random Forest with default parameters is the best performer for bagging method.
  - Monthlyincome is the most important feature for prediction followed by DurationOfPitch, Passport_1 and NumberOfTrips.
  - Accuracy on training set : 1.0
  - Accuracy on test set : 0.929107021131561
  - Recall on training set : 1.0
  - Recall on test set : 0.7028985507246377
  - Precision on training set : 1.0
  - Precision on test set : 0.8981481481481481

- Tuned XGBoost model is the best performer mode. It has high performance metrics, and has the best recall values.
  - Designation_Executive is the most important feature for prediction followed by Passport_1, MaritalStatus_Single and CityTier_3,
  - Accuracy on training set : 0.9929845074539608
  - Accuracy on test set : 0.9222903885480572
  - Recall on training set : 0.9782608695652174
  - Recall on test set : 0.7391304347826086
  - Precision on training set : 0.984375
  - Precision on test set : 0.8292682926829268

Recommendations

- Company can focus on targeting customer with these strong important features:
  a. Designation_Executive
  b. Passport_1
  c. MaritalStatus_Single
  d. CityTier_3

- In both of the strongest models, Tuned Random Forest and Tuned XGBoost, a common important feature is Passport_1. The company can focus on targeting customer that have a passport.

greatlearning
Power Ahead

Happy Learning !

Esteban Ordenes

Post Graduate Program in Data Science and Business Analytics

PGP-DSBA-UTA-Dec20-A